Incorporating IMERG Satellite Precipitation Uncertainty into Seasonal and Peak Streamflow Predictions using the Hillslope Link Hydrological Model

Samantha H. Hartke[1], Daniel B. Wright[1], Felipe Quintero[2], Aline S. Falck[3]

[1]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, Wisconsin, USA.

[2]IIHR—Hydroscience & Engineering, Iowa Flood Center, University of Iowa, Iowa City, USA

[3]National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), São José dos Campos, SP, Brazil

Abstract

In global applications and data sparse regions, which comprise most of the earth, hydrologic model-based flood monitoring relies on precipitation data from satellite multisensor precipitation products or numerical weather forecasts. However, these products often exhibit substantial errors during the meteorological conditions that lead to flooding, including extreme rainfall. The propagation of precipitation forcing errors to predicted runoff and streamflow is scale-dependent and requires an understanding of the autocorrelation structure of precipitation errors, since error autocorrelation impacts the accumulation of precipitation errors over space and time in hydrologic models. Previous efforts to account for satellite precipitation uncertainty in hydrologic models have demonstrated the potential for improving streamflow estimates; however, these efforts use satellite precipitation error models that rely heavily on ground reference data such as rain gages or weather radar and do not characterize the nonstationarity of precipitation error autocorrelation structures. This work evaluates a new method, the Space-Time Rainfall Error and Autocorrelation Model (STREAM), which stochastically generates possible true precipitation fields, as input to the Hillslope Link Model to generate ensemble streamflow estimates. Unlike previous error models, STREAM represents the nonstationary and anisotropic autocorrelation structure of satellite

precipitation error and does not use any ground reference to do so. Ensemble streamflow predictions are compared with streamflow generated using satellite precipitation fields as well as a radar-gage precipitation dataset during peak flow events. Results demonstrate that this approach to accounting for precipitation uncertainty effectively characterizes the uncertainty in streamflow estimates and reduces the error of predicted streamflow. Streamflow ensembles forced by STREAM improve streamflow prediction nearly to the level obtained using ground-reference forcing data across basin sizes.

Keywords: satellite methods, precipitation, catchment dynamics, uncertainty

1. Introduction

1.1 Motivation

Hydrologic models are central to efforts to mitigate the devastating impacts of floods and are used around the world to predict streamflow, with particular interest in high streamflow events that cause flooding (e.g., Alfieri et al., 2013; Wu et al., 2014). Despite continual advances, these models cannot perfectly predict streamflow for a number of reasons. In addition to model uncertainty (i.e. imperfect representation of physical processes and parameter uncertainty), forcing uncertainty due to errors in rainfall data—whether obtained from gauge data, radar, numerical weather prediction model (NWP) forecasts, or satellite multi-sensor precipitation (SMP) products—leads to errors in runoff and streamflow estimates (Hong et al., 2006; Sperna Weiland et al., 2015). In global hydrologic modeling studies, precipitation forcing data has been found to be the primary driver of predictive uncertainty, and using parameter calibration to compensate for biases in meteorological forcing data can actually result in inconsistent performance between calibration and evaluation periods, leading to calls for "improved characterization of global rainfall

amounts at spatial resolutions of 0.5° and smaller" (Fekete et al., 2004; Sperna Weiland et al., 2015).

Precipitation uncertainty in basins in high latitudes and the tropics leads to hydrologic uncertainty of the same or greater magnitudes due to the relatively moist soil conditions and propensity of rainfall to become runoff in these regions (Biemans et al., 2009; Fekete et al., 2004). Precipitation error propagation is more variable in semi-arid regions where the rainfall-runoff generation processes are highly nonlinear (Fekete et al., 2004). Initial and simulated soil moisture conditions in watersheds, which are also critical to accurate streamflow prediction, are dependent on multiple previous precipitation events (Alvarez-Garreton et al., 2014; Tramblay et al., 2010, 2011).

The issue of precipitation uncertainty in streamflow prediction is particularly prevalent in regional-to-global-scale efforts, which due to lack of alternatives, typically must rely on either precipitation from satellites or numerical weather models—both of which are prone to substantially larger errors than in-situ gage measurements. Examples include the Global Flood Monitoring System (GFMS; Wu et al., 2014) and the Global Flood Awareness System (GloFAS; Alfieri et al., 2013). These use NASA's Integrated MultisatellitE Retrieval for the Global Measurement Mission (IMERG) and the ECMWF Integrated Forecast System (EFS) data, respectively, to generate global streamflow predictions. Similar to other satellite multi-sensor precipitation products, IMERG precipitation estimates exhibit considerable systematic bias and random error, and IMERG often overestimates rainfall during extreme precipitation events (e.g., Anjum et al., 2018; Asong et al., 2017; Gilewski & Nawalany, 2018; Omranian et al., 2018; Wang et al., 2017). A number of studies have demonstrated that IMERG can introduce large uncertainties into streamflow predictions when compared with gauge-based or radar-gauge products, although

71 the extent to which IMERG-based results underperform varies (Amorim et al., 2020; Jiang &
72 Bauer-Gottwein, 2019; N. Li et al., 2016). For instance, IMERG-forced streamflow predictions
73 across China were shown to exhibit substantially more error in arid watersheds than humid
74 watersheds (Jiang et al., 2019).

75 For a thorough review of the use of satellite multi-sensor precipitation (SMP) products in
76 hydrologic modeling, see Quintero et al. (2016). Hydrologic modeling is not the only satellite
77 precipitation application that contends with satellite precipitation uncertainty; land surface
78 modeling, snow simulations, groundwater modeling, and landslide hazard assessment are also
79 susceptible to errors in output when ingesting erroneous precipitation data (e.g., Hartke et al., 2020;
80 Maggioni et al., 2011; Marc et al., 2022; Pradhan & Indu, 2021; Raleigh et al., 2015; Schreiner-
81 McGraw & Ajami, 2020; Serpetzoglou et al., 2010; Shrestha et al., 2020). Although this work
82 focuses on accounting for uncertainty in SMP data, numerical weather prediction model (NWP)
83 precipitation fields also exhibit similar levels of bias and random error, including high
84 uncertainties during extreme events, due to a highly dynamic atmosphere and the sensitivity of
85 model precipitation forecasts to a range of model parameters (e.g., Lowrey & Yang, 2008; Luitel
86 et al., 2018; Moosavi et al., 2021; Nasrollahi et al., 2012; Nogueira, 2020).

87 ## 1.2 Precipitation Error Modeling Background

88 Within hydrologic models runoff is accumulated over a range of basin sizes, and precipitation
89 error propagation is dependent on both basin scale as well as the autocorrelation of precipitation
90 errors in space and time (Cunha et al., 2012; Nijssen & Lettenmaier, 2004; Nikolopoulos et al.,
91 2010). Error autocorrelation describes the intuitive fact that an overestimation of precipitation by
92 IMERG at a pixel in a given timestep likely corresponds to an overestimation by IMERG in

surrounding pixels and timesteps. While accumulation during rainfall-runoff generation processes can serve to "average out" random errors in precipitation datasets, it can also serve to accumulate a field of precipitation with correlated errors (i.e. precipitation being overestimated over an entire storm system) and propagate these into streamflow predictions. Cunha et al. (2012) found "the efficiency of the river basin in filtering out random errors to be highly sensitive to the presence of spatial correlation in errors," and that "when rainfall errors are correlated in space, the process of aggregation and attenuation by the river network is not as effective in filtering out uncertainties." Correctly simulating the autocorrelation of precipitation error fields is paramount to capturing precipitation uncertainty at coarser scales (e.g., basin scales; Hartke et al., 2022). Just like the correlation structure of precipitation, the correlation structure of precipitation error fields is anisotropic and nonstationary in time and space, changing depending on the observed precipitation system and retrieval conditions (i.e., errors may exhibit lower spatial autocorrelation during a scattered precipitation event than during a larger—and more highly spatially autocorrelated—frontal precipitation event).

A number of SMP error models have been introduced that generate distributions to describe the uncertainty surrounding a given SMP estimate at a single time and pixel (i.e., Kirstetter et al., 2018; Maggioni et al., 2014; Wright et al., 2017; see Section 3.3). One of the major limitations of these "pixel-scale" error formulations is that they do not relate the uncertainty of SMP estimates across space and time; there is no intuitive way to combine the uncertainty distributions at every pixel in a field to create an ensemble of precipitation error fields. Previous work has attempted to link pixel-scale precipitation uncertainty estimates using calibrated correlation coefficients or lengths to quantify the space-time autocorrelation structure of errors (e.g., Ciach et al., 2007; Hossain & Anagnostou, 2006), but such approaches do not represent the nonstationarity of satellite

precipitation errors and rely on ground reference data for calibration. Previous studies on ensemble SMP error correction in hydrologic modeling have utilized the Two-Dimensional Satellite Rainfall Error Model (SREM2D; Hossain & Anagnostou, 2006), which does not account for the nonstationarity and anisotropy of SMP error correlation structures (Falck et al., 2015, 2018, 2021; Maggioni et al., 2013). These studies demonstrate that the inclusion of precipitation uncertainty using ensemble methods does generally improve streamflow prediction, although not always in large basins when an error model applies bias correction to entire basin areas (as shown with SREM2D in Falck et al., 2015), but do so using an approach that is not viable for most of the world, due to a scarcity of ground-based data which is unlikely to change.

1.3 Incorporating satellite precipitation uncertainty into hydrologic modeling

Accounting for precipitation uncertainty in hydrologic modeling applications poses a challenge because of the probabilistic nature of precipitation uncertainty, the nonstationary correlation structure of precipitation errors, and the need to calibrate precipitation uncertainty models with little to no ground-reference data in many parts of the world. The Space-Time Rainfall Error and Autocorrelation Model (STREAM) was developed with an eye to overcoming these challenges (Hartke et al. 2022). STREAM stochastically generates a precipitation ensemble that represents the possible true precipitation based on a satellite precipitation field (or a numerical weather forecast). It consists of two pieces: 1) using an uncalibrated, nonstationary error autocorrelation scheme that requires no ground reference data and 2) pixel-scale uncertainty estimates that can be obtained from existing error model formulations. By not calibrating the autocorrelation scheme on historical ground reference or SMP error fields and instead replicating the changing, anisotropic autocorrelation structure of SMP fields, STREAM nimbly incorporates nonstationarity into its representation of SMP uncertainty. Like previous ensemble-based approaches, STREAM

ensembles can be ingested by SMP applications, including hydrologic models, without any modification to the model structure.

One strategy to address precipitation uncertainty in hydrologic modeling is bias correction (e.g., Charles et al., 2020; Ciupak et al., 2019; Habib et al., 2014; Ji et al., 2020). However, when applied in near real-time this approach only adjusts for systematic bias and does not consider or account for the substantial random error in SMP and NWP datasets. This work employs a bias correction scheme in order to compare such an approach with one that considers the full range of precipitation uncertainty – both systematic bias and random error.

The goal of this work is to understand how precipitation uncertainty impacts streamflow estimates, with an emphasis on peak streamflow events, and assess whether STREAM presents a suitable way to account for IMERG uncertainty and improve predicted streamflow. In this work, IMERG uncertainty is incorporated into the Iowa Flood Center's distributed hydrologic model, the Hillslope Link Model (HLM), using precipitation ensemble fields generated by STREAM. The resulting ensemble of streamflow estimates are compared against USGS observations as well as streamflow estimates from HLM when forced with IMERG-Early, a bias-corrected version of IMERG-Early, and a ground-reference product, NEXRAD Stage IV (Lin, 2011). Unlike previous attempts to account for SMP uncertainty, namely by using SREM2D, STREAM ensembles model the nonstationarity of SMP error correlation structures and do not rely on ground-reference precipitation data to do so.

This paper is structured as follows: the study area in Iowa and precipitation and streamflow datasets are described in Section 2. Section 3 details the methodology of the Hillslope Link Model and STREAM and presents the chosen evaluation metrics for streamflow results. Section 4

161 presents HLM simulation results using a range of precipitation inputs. Results and implications of
162 this work are discussed in Section 5.

163 2. Study Area and Data

164 2.1 Study Area and Period

165 The study area comprises the state of Iowa (roughly 97°W to 90°W, 40°N to 44°N), including
166 the major river basins of the Cedar, Iowa, Skunk, and Des Moines rivers that drain into the
167 Mississippi River along the eastern border of the state (Figure 1c). This study area is classified as
168 humid continental on the Koppen climate scale, covers a range of watersheds and sub watersheds,
169 and is comparable to or larger than many study areas used to validate hydrologic performance of
170 SMP data (e.g., Amorim et al., 2020; Maggioni et al., 2013; Nikolopoulos et al., 2010). In the past
171 three decades, the state has experienced several widespread flooding events with substantial social
172 and economic tolls, with the most notable events occurring in 1993, 2008, and 2011 (Lott, 1993;
173 Mutel, 2010; Vennapusa & White, 2015). The study period covers 2008 – 2013, which includes
174 one of the largest flood events that Iowa has experienced in June 2008 (Budikova et al., 2010;
175 Smith et al., 2013).

176 2.2 Precipitation Data

177 The NEXRAD Stage IV radar-gauge product, available over the Continental U.S. (CONUS)
178 at an hourly, 1/24˚ resolution (Lin, 2011), is used as the ground-reference precipitation product in
179 this work. Although Stage IV is not exempt from errors, it has been considered to exhibit negligible
180 uncertainty relative to SMP data in previous SMP studies (e.g., Aghakouchak et al., 2011), and is
181 considered highly accurate in this study region (Quintero et al., 2020b).

NASA's IMERG satellite product is available globally from 2000 to present at a half-hourly, 0.1° resolution in three versions: IMERG-Early at a 4-hour latency, available in near real-time but excluding remote sensing data following a satellite pass, IMERG-Late at a 12-hour latency, which incorporates additional remote sensing retrievals, and IMERG-Final at an approximately 2.5 month latency, which assimilates gauge data to improve product accuracy (Huffman et al., 2019; Tan et al., 2019). In this work, IMERG-Early is used because of its low latency and availability for early warning systems. IMERG-Early is rescaled to an hourly resolution to match the temporal resolution of Stage IV data during calibration of the SMP error model required for STREAM (Section 3.2).

The threshold of detection for all precipitation data in this work is 0.1 mm/hour, consistent with previous sub daily precipitation studies (e.g., Germann & Zawadzki, 2002; Li et al., 2021).

2.3 Streamflow Data

USGS gauge data throughout the state of Iowa was used to validate the streamflow estimated by the Hillslope Link Model (Section 3.1). A total of 192 USGS stations are used. Periods with missing streamflow data are excluded from analysis. Fourteen gauge sites which are heavily influenced by reservoir operations or other anthropogenic diversions are also excluded from analysis since the hydrologic model described in Section 3.1 does not account for these effects. Roughly 20% of the resulting 178 gauge sites have an upstream area less than 100 km^2; 30% of sites have upstream areas between 100 and 1,000 km^2; 15% of gauge sites have upstream areas greater than 5,000 km^2 with the largest upstream area being 32,645 km^2 at the Des Moines River at Keosauqua (Figure 1d).

3. Methods

3.1 Hillslope Link Model

The Iowa Flood Center's Hillslope Link Model is a continuous rainfall-runoff model that routes runoff calculated at individual 'hillslopes' (Figure 1b) into links which are connected via channel routing (Figure 1a; Krajewski et al., 2017). The HLM was developed initially as part of the Iowa Flood Information System (IFIS; Demir et al., 2018; Demir & Krajewski, 2013; Krajewski et al., 2017) in response to the devastating 2008 flooding in eastern Iowa and began operational use in 2012, providing distributed streamflow predictions and a flood potential index (Quintero et al., 2020a) every 15 minutes at sites across the state (Figure 1c). This study uses the same parameters as in the operational model and does not calibrate the HLM model for specific precipitation datasets. Although the HLM is not a continental or global scale model that utilizes SMP or NWP data, it shares the objective of other operational forecast models to provide expedient, accurate data to prepare communities for potential high flow and flooding events.

In the HLM, the landscape is divided into individual channels and hillslopes based on USGS digital elevation model (DEM) data with a 90-meter resolution. Rainfall input is partitioned into soil moisture, drainage, and runoff fluxes. The uncertainty in rainfall input data is viewed as one of the largest uncertainty sources for model predictions (Quintero et al., 2020b).

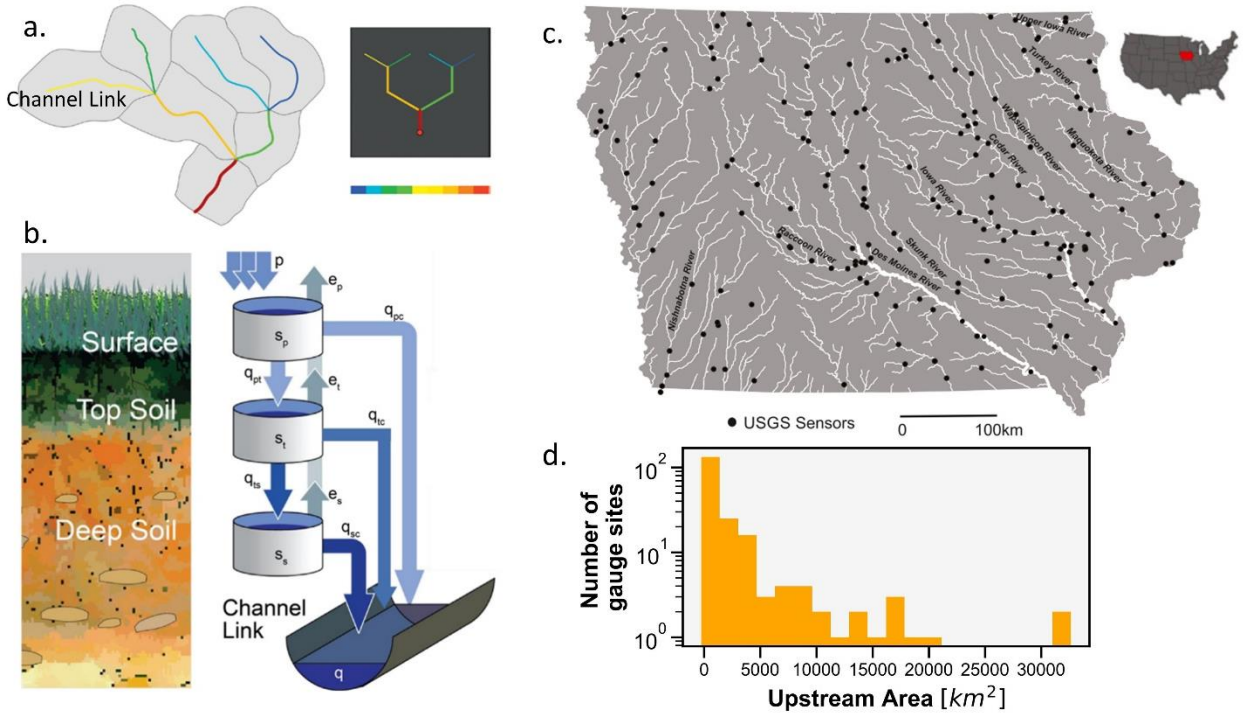


Figure 1. Methodology schematic of the Iowa Flood Center's Hillslope Link Model. (a) A channel routing program combines the (b) runoff calculated from hillslopes for individual channel links. (c) Model streamflow estimates are output and compared to observed streamflow at USGS gauge sites across the state of Iowa. (d) Histogram of upstream area at all sites. Figures from Quintero et al. (2020a) and Quintero et al. (2020b).

3.2 Space-Time Rainfall Error and Autocorrelation Model (STREAM)

The Space-Time Rainfall Error and Autocorrelation Model (STREAM) was developed to allow pixel-scale SMP uncertainty estimates to be combined in space and time by simulating the space-time autocorrelation structure of SMP error using minimal ground reference data (Hartke, Wright, Li, Maggioni, & Dalia, 2021). The space-time correlation structure of an SMP field – in this application, IMERG – is adopted as the space-time correlation structure of SMP error at each time step using the python package pySTEPs and a semi-Lagrangian advection scheme (Pulkkinen et al., 2019; Figure 2; see Hartke et al., (2022) for further details). IMERG uncertainty at each pixel and timestep is represented by a distribution that is conditional on an IMERG precipitation estimate; selecting any value randomly from this conditional distribution produces equally

probable values for the true precipitation. To generate a field of possible true precipitation values, however, requires consideration of the correlation between satellite precipitation errors. Error autocorrelation describes the intuitive fact that an overestimation of precipitation by IMERG at a pixel in a given timestep likely corresponds to an overestimation by IMERG in surrounding pixels and timesteps. By stochastically generating uniform noise fields that replicate the autocorrelation structure of IMERG fields, a proxy for the unknown autocorrelation structure of the IMERG error field, possible true precipitation values can be selected from conditional distributions to generate precipitation fields which incorporate autocorrelated precipitation errors (Figure 2). The Censored

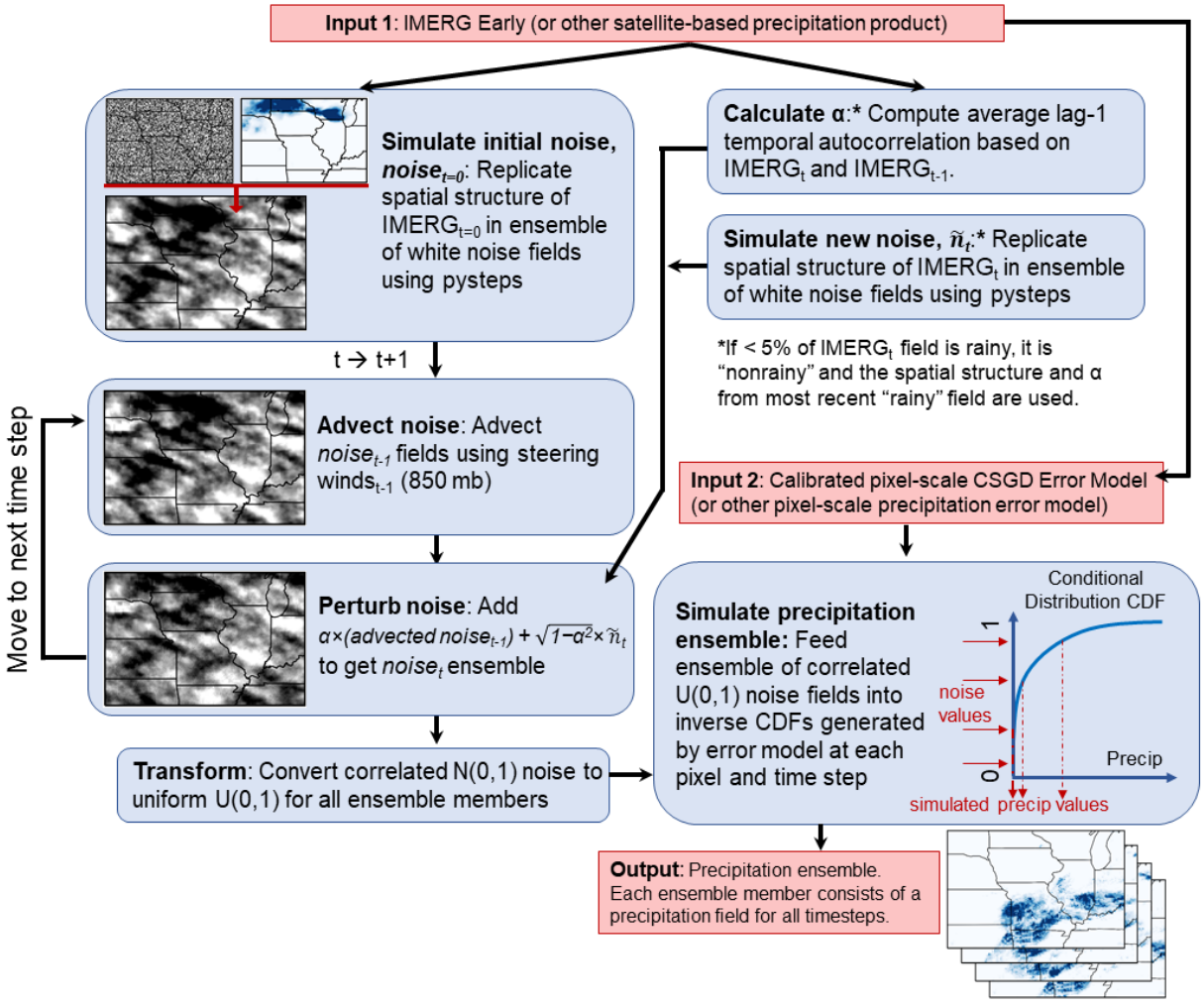


Figure 2. STREAM schematic (Hartke et al. 2022)

Shifted Gamma Distribution (CSGD) error model and the conditional distributions it generates are further described in Section 3.3.

As in Hartke et al. (2022), the CSGD error model is used here to generate pixel-scale uncertainty distributions for IMERG estimates at every time step and pixel in the study area. For further information on STREAM, see Hartke et al. (2022).

3.3 The Censored Shifted Gamma Distribution Error Model and a Bias-corrected IMERG

The Censored Shifted Gamma Distribution error model was first introduced by Scheuerer & Hamill (2015) for post-processing of NWP precipitation fields and was adapted by Wright et al. (2017) to characterize the uncertainty of single satellite precipitation estimates across the continental U.S. The CSGD which is used to describe the uncertainty around a precipitation estimate is an adaptation of the two-parameter gamma distribution with an additional "shift" parameter δ that shifts the probability density function (PDF) leftward. The density left of zero represents the probability of zero precipitation, while the density at any value greater than zero represents the likelihood of that amount of precipitation (Figure 3c). The reparameterized distribution is then left-censored at zero, replacing all negative values with zero.

A regression model, described by error model parameters α_1, α_2, α_3, and α_4, is trained based on contemporaneous co-located SMP and ground-truth observations. The calibrated model generates unique "conditional" CSGD parameters μ(t), σ(t), and δ(t) at any time *t* as a function of the model parameters and the SMP estimate $R_s(t)$:

$$\mu(t) = \frac{\mu_c}{\alpha_1}\log\left\{1 + (\exp(\alpha_1) - 1)\left[\alpha_2 + \alpha_3 \frac{R_s(t)}{\bar{R}}\right]\right\} \quad \text{Equation 1,}$$

$$\sigma(t) = \alpha_4 \sigma_c \sqrt{\frac{\mu(t)}{\mu_c}} \quad \text{Equation 2,}$$

$$\delta(t) = \delta_c \quad \text{Equation 3,}$$

Where μ_c, σ_c, and δ_c are the parameters of the climatological CSGD, calculated using SMP data (Figure 3a and b), and $\bar{R}$ is the mean of SMP estimates. μ(t), σ(t), and δ(t) are the parameters of a conditional CSGD describing the uncertainty around SMP estimate $R_s(t)$ (Figure 3c). Equations 1-3 describing the CSGD are written in terms of its mean and standard deviation but can be reparametrized in terms of shape and scale parameters. The error model parameters and climatological CSGD can be trained for a single location using ground reference and SMP timeseries from a single grid cell or can be regionalized by concatenating the timeseries from multiple grid cells and using the resulting timeseries during calibration. In this application of the STREAM (Section 3.2), the CSGD error model is trained regionally over 25 0.1° grid cells at a time (5 grid cells wide and 5 grid cells high; Figure 3a). This approach ensures that sufficient data is used during model training and that differences in IMERG error characteristics are represented across the study region.

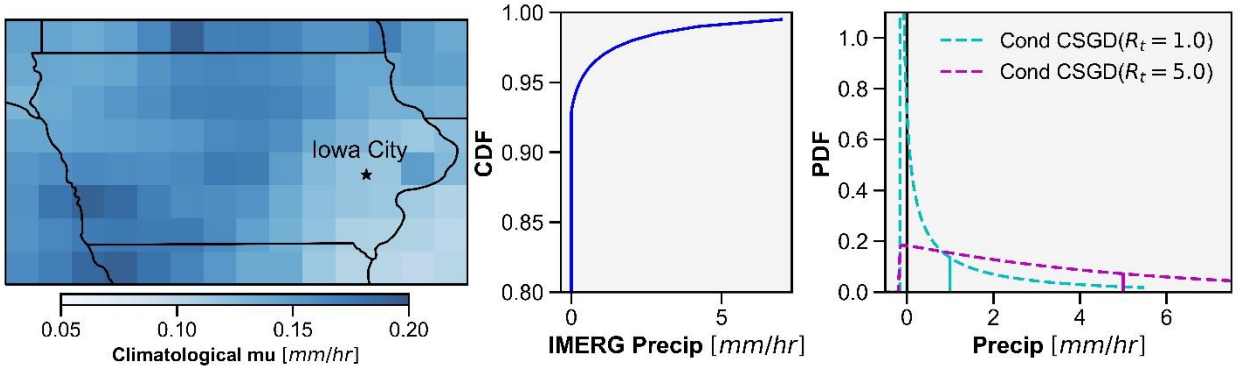


Figure 3. (Left) Climatological mu of 5 by 5 pixel "regional" CGSD error models in study area. (Middle) CDF of climatological CSGD for region over Iowa City. (Right) Conditional CSGDs generated using regional error model calibrated over Iowa City for IMERG estimates of 1 mm/hr (blue) and 5 mm/hr (pink).

An additional precipitation dataset is generated by selecting the mean value of the conditional distribution generated for each IMERG estimate in space and time, adjusting the IMERG estimate for systematic bias. This precipitation dataset is effectively a bias-corrected version of IMERG and is referred to as bias-corrected IMERG in the remainder of this work.

3.4 Experimental Set Up

The HLM is run for April – October in 2008 – 2013 at a 1-hour timestep with the following five precipitation inputs: 0.1° IMERG, 0.1° Stage IV, 1/24° Stage IV, 0.1° STREAM ensemble fields (Section 3.2), and a 0.1° bias-corrected IMERG field (Section 3.3). All precipitation inputs produce deterministic streamflow estimates except the STREAM ensemble fields, which produce an ensemble of streamflow estimates. Two resolutions of Stage IV precipitation are used so that the coarser 0.1° product, bilinearly interpolated from the native 1/24° resolution product, matches the resolution of IMERG, and any differences between the performance of the two Stage IV products can be attributed to the resolution of the precipitation data. The first month of the HLM simulation in each year (April) is discarded as a spin-up period, and May – October is used for evaluation.

3.5 Streamflow Peak Event Identification

In order to assess HLM streamflow performance during high flow events, we identify peak events using the wavelet-based event identification method proposed in Towler & McCreight (2021) and Liu et al. (2011). This method assesses the statistical significance of the spectral power "jumps" that large streamflow events represent when the entire streamflow record is transformed into spectral space using a wavelet transform, which creates a representation of the signal in both the time and frequency domain (Figure 4 a-c). We assessed the power spectrum averaged over the

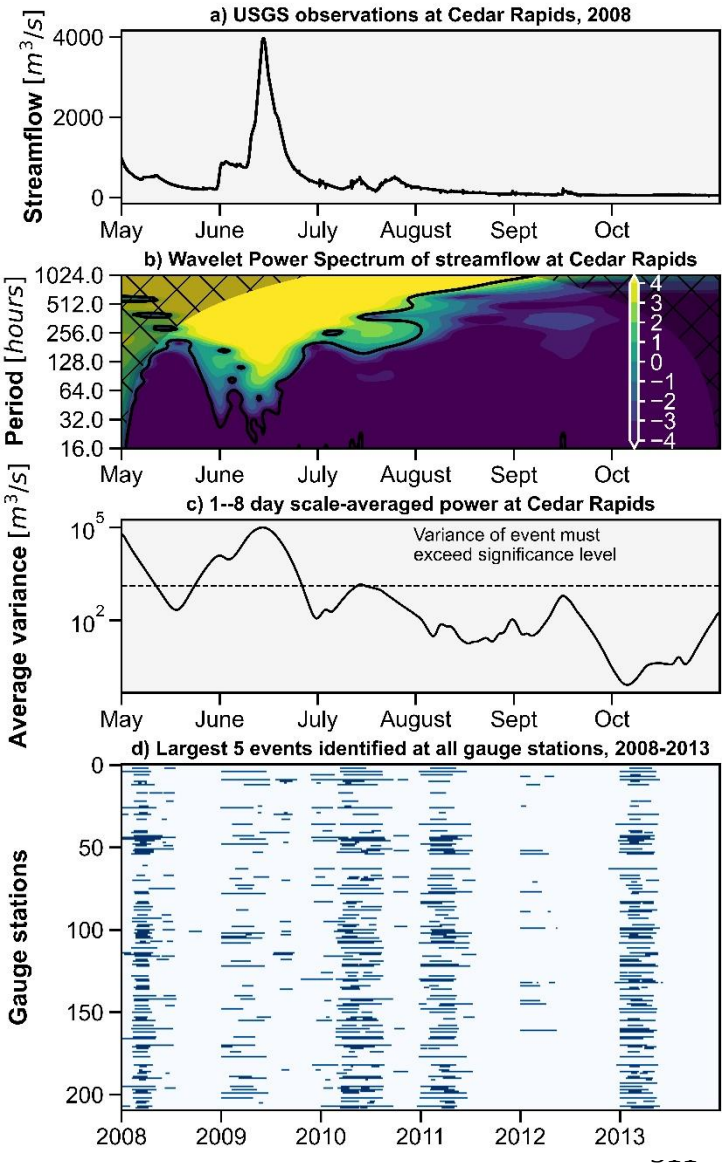


Figure 4. (a) Hydrograph, (b) Wavelet power spectrum of streamflow, and (c) Variance of the power spectrum averaged over 1-8 day (24-192 hour) period for 2008 at the Cedar River at Cedar Rapids. (d) The identified five largest events at each gauge site over 2008-2013 study period.

1-8 day period to find statistically significant events at this scale. Wavelet-based event identification is advantageous because it allows comparison of peak event characteristics (peak flow, flow volume, event duration) without the complication of timing errors.

Intermittent periods with missing data in USGS gauge records are linearly interpolated for the purposes of event identification and calculation of total event volume. After identifying events at each gauge site using observed USGS streamflow (Figure 4d), the peak event flow and total event volume are calculated for the five largest events, as determined by event volume, using the observed streamflow and the simulated streamflow obtained by forcing HLM with all precipitation datasets. At almost every gauge site, the top five identified events included events during June 2008 and June 2013. The mean absolute error (MAE) of IMERG and Stage IV simulations and mean continuous ranked probability score (MCRPS;

described in Section 3.6) of STREAM simulations are calculated for peak flow and event volume at these events.

3.6 Performance metrics

The normalized root mean squared error (NRMSE), also referred to as the relative RMSE (Falck et al., 2015), is calculated for IMERG-based and Stage IV-based streamflow as well as for the mean of the STREAM-based streamflow ensemble.

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{Q}_t - Q_t)^2}}{\frac{1}{N}\sum_{t=1}^{N} Q_t} \quad \text{Equation 4,}$$

Where Q_t is the observed streamflow and $\hat{Q}_t$ is the predicted streamflow at timestep *t*. The mean absolute error (MAE) for deterministic streamflow estimates and the mean continuous ranked probability score (MCRPS) for streamflow ensembles are calculated. The continuous ranked probability score of an ensemble streamflow estimate is analogous to the absolute error of a single streamflow estimate and reduces to the absolute error as the ensemble size decreases to one; therefore, the MCRPS of a probabilistic streamflow timeseries is analogous to the MAE of a deterministic timeseries, allowing for comparison of performance between probabilistic and deterministic streamflow predictions.

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|\hat{Q}_t - Q_t\right| \quad \text{Equation 5,}$$

$$MCRPS = \frac{1}{N}\sum_{t=1}^{N}\int_{\mathbb{R}}\left[F\left(\hat{Q}_t\right) - I(\hat{Q}_t \geq Q_t)\right]^2 d\hat{Q}_t \quad \text{Equation 6,}$$

Where $F(\hat{Q}_t)$ is the forecasted distribution of streamflow, in this case an ensemble rather than an explicit distribution, and $I[\bullet]$ is an indicator function. The ideal value for NRMSE, MAE, and MCRPS is 0, and higher values indicate worse performance.

The containing ratio (CR) is the proportion of observed streamflow data which lies within the range of the streamflow ensemble generated using STREAM precipitation ensembles. This probabilistic performance metric, and its counterpart, the exceedance ratio (ER), have been used by previous work in the satellite precipitation, hydrologic modeling, and forecast verification communities to assess ensemble performance (e.g. Franz & Hogue, 2011; Hartke et al., 2021; Maggioni et al., 2013; Xiong & O'Connor, 2008)

$$CR = \frac{1}{n}\sum_{t=1}^{n} I[Q_t] \quad \text{Equation 7,}$$

Where the indicator function $I[\bullet]$ is equal to one when Q_t is bracketed by the minimum and maximum values of the predicted streamflow ensemble at timestep t and is equal to zero when Q_t lies outside the ensemble. The ideal value for CR is 1, indicating that the ensemble brackets observed streamflow in every instance, while the poorest performance would be indicated by a value of 0.

The uncertainty ratio (UR) is the ratio between ensemble spread and the corresponding observed streamflow at each timestep and has been used in previous studies of probabilistic streamflow and soil moisture (Falck et al., 2015; Maggioni et al., 2011).

$$UR = \frac{\sum_{t=1}^{n}(\hat{Q}_{t_{max}} - \hat{Q}_{t_{min}})}{\sum_{t=1}^{n} Q_t} \quad \text{Equation 8,}$$

The UR can vary between 0 and infinity, with lower values indicating lower ensemble spread relative to observed streamflow. A high CR and low UR are ideal for ensemble predictions because

this indicates that an ensemble is consistently bracketing observations while maintaining a small ensemble spread relative to observed streamflow values. Periods with missing data in USGS gauge records are excluded when calculating the above performance metrics.

4. Results

4.1 Performance of HLM streamflow estimates

Figure 5 shows the containing ratio (CR) of the streamflow ensemble generated using STREAM ensemble precipitation input and the percent reduction in MAE of the IMERG-based streamflow when compared to the MCRPS of the STREAM-based streamflow ensemble. The CR of the STREAM-based streamflow ensemble varies across the study area but is consistently high in watersheds in eastern and western Iowa. Lower CR values in the Des Moines Lobe landform in central Iowa correspond to an area where the HLM configuration has low performance, mainly because the model does not account for changes in infiltration and flow due to agricultural tile drainage systems (Quintero et al., 2020). The MCRPS of the STREAM-based streamflow

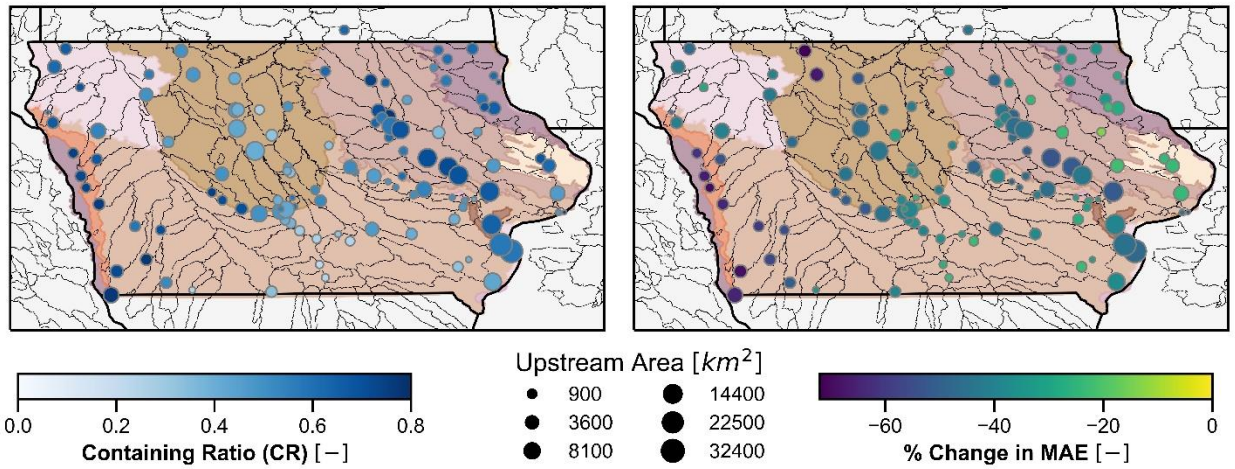


Figure 5. (Left) Map showing containing ratio (CR) of streamflow ensemble generated by HLM using STREAM ensemble as input and (Right) Percent difference between MAE of IMERG-based streamflow and MCRPS of STREAM-based streamflow ensemble. The STREAM-based streamflow ensemble has a MCRPS that is 20 - 60% lower than the MAE of IMERG-based streamflow at most gauge sites in the study area. Different geomorphological landforms across the state are shaded.

ensemble is lower than the MAE of IMERG-based streamflow across all watersheds, reducing the MAE by up to 80% in some locations.

Figure 6 plots the NRMSE, MAE and MCRPS of IMERG-, Stage IV-, and STREAM-based streamflow estimates. The CR and UR of the STREAM-based ensemble are also presented in panels c and d. Stage IV-based streamflow generally exhibits the lowest error; however, the STREAM-based streamflow exhibits errors nearly as low as that of Stage IV-based streamflow and much lower than that of the IMERG-based simulations. MAE and MCRPS increase with upstream area, which is expected since streamflow magnitudes are greater in bigger basins. The reduction in MCRPS relative to MAE by STREAM-based streamflow estimates increases substantially with basin size; the STREAM-based streamflow ensemble exhibits a MCRPS 50% lower than the MAE of IMERG-based streamflow estimates at gauge sites with upstream areas greater than 10,000 km^2. Except for basins < 100 km^2, the CR of STREAM simulations generally increases with upstream area when basins < 100 km^2 (the size of a single IMERG pixel) are excluded, while the UR of the STREAM ensemble consistently decreases with upstream area. The

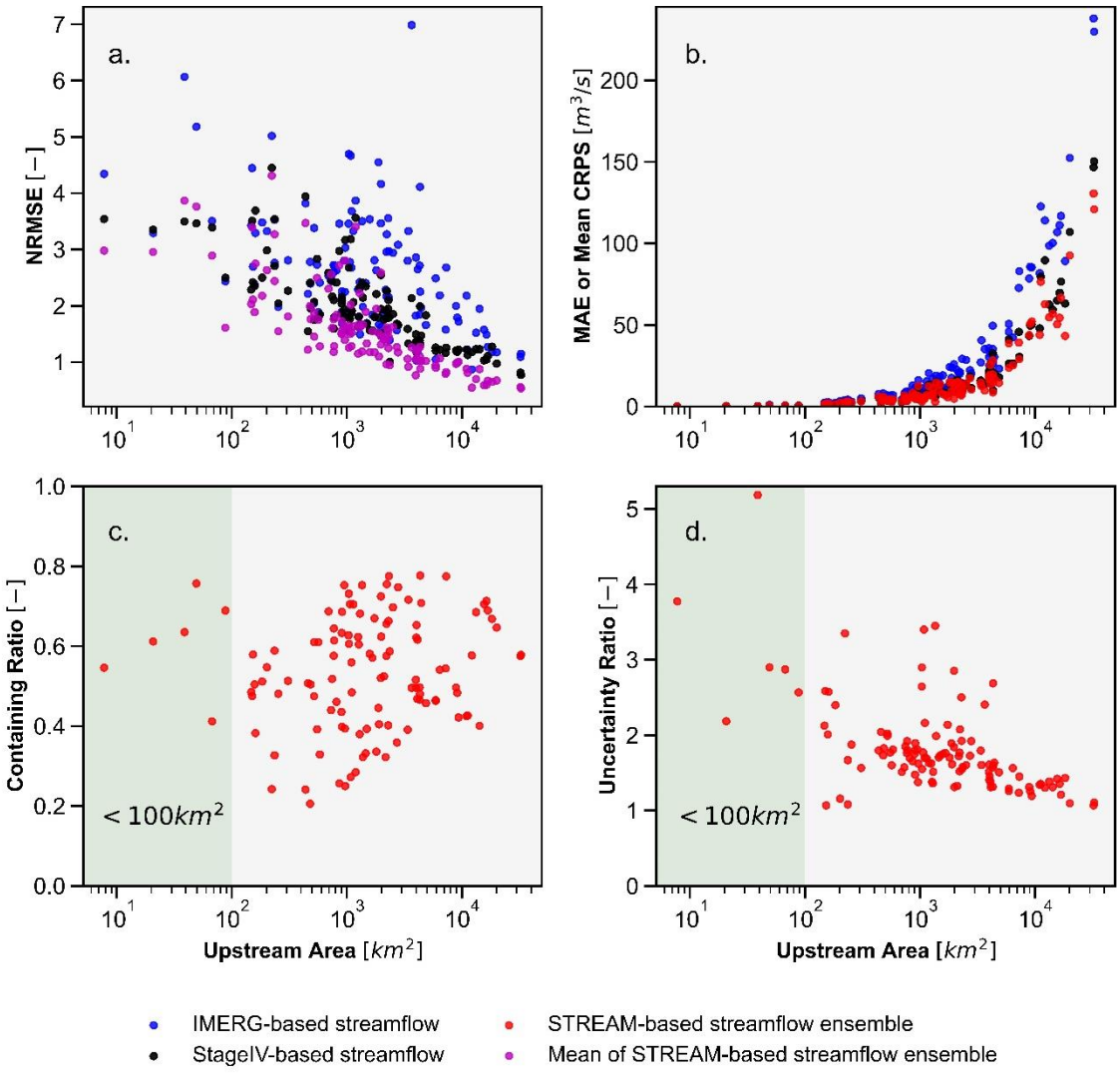


Figure 6. (a) NRMSE and (b) MAE and MCRPS calculated for all streamflow estimates. (c) Containing Ratio, and (d) Uncertainty Ratio calculated for STREAM-based streamflow estimates April-October 2008-2013 at USGS gauge sites across Iowa as a function of upstream area. The green areas in (c) and (d) highlight results from gauges with upstream areas less than 100 km^2, an area smaller than that of a single IMERG grid cell.

384 large ensemble spreads in small basins reflected by high UR may be due to the limitations of
385 hydrological models in simulating fast hydrological responses in small basins, when remote
386 sensing products such as IMERG are too coarse to capture (Nimmo et al., 2021; Széles et al.,
387 2020).

Figure 7 displays the MCRPS and CR of STREAM-based streamflow ensembles as a function of ensemble size and upstream area. Ensemble members were chosen at random to generate this plot. MCRPS generally decreases and CR generally increases with increasing ensemble size. As shown previously in Figure 5, MCRPS increases with upstream area. MCRPS decreases with ensemble size in basins less than 4500 km^2 in size until an ensemble size of 10 is reached, after which the MCRPS appears to stabilize. In basins larger than 4500 km^2, a larger ensemble appears to improve MCRPS until an ensemble size of 20 is reached. The CR demonstrates greater change with ensemble size; for all basin sizes, the CR continuously increases with ensemble size, and basins less than 100 km^2 (greater than 2000 km^2) in size eventually level out at a CR of approximately 0.6 (0.55). The upstream area range that consistently has the lowest CR is 100 – 750 km^2.

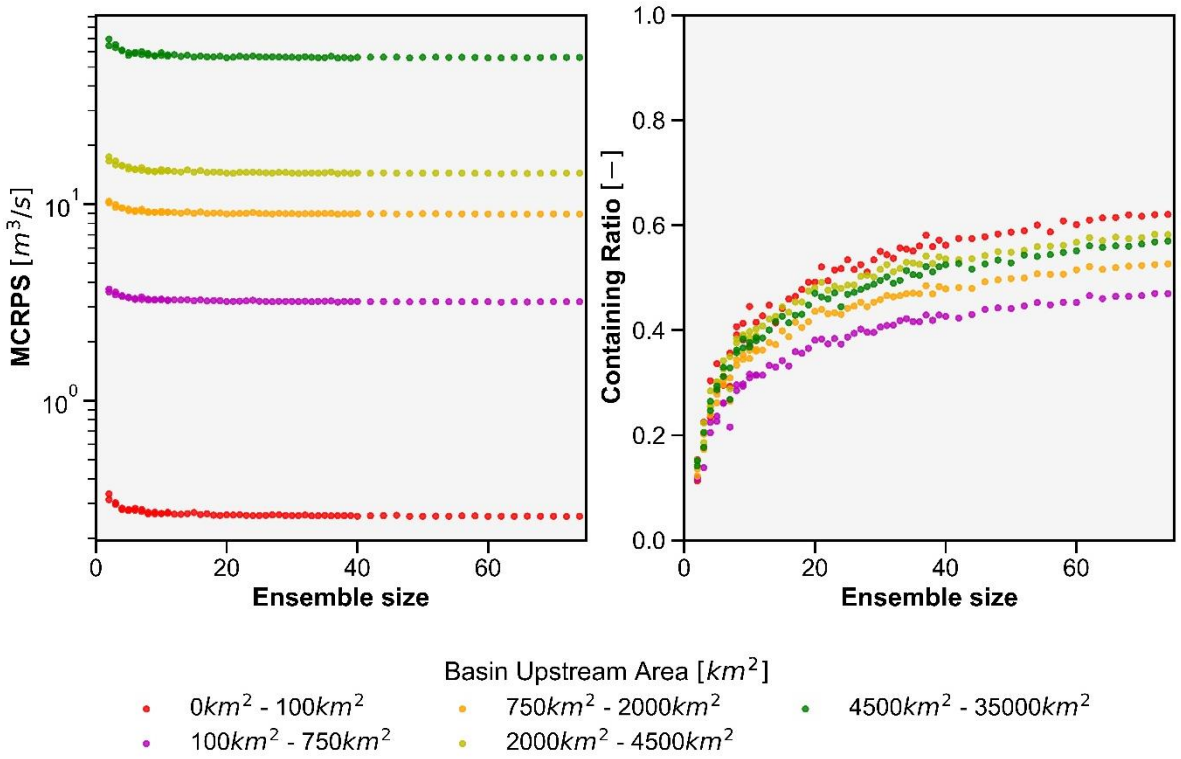


Figure 7. (Left) MCRPS and (Right) Containing Ratio of STREAM-based streamflow ensemble as a function of ensemble size and basin area. The MCRPS decreases with increasing ensemble size and is lowest in basins with the smallest upstream areas. CR increases with ensemble size.

4.2 Performance of HLM streamflow estimates during extreme flow events

Figure 8 illustrates observed and simulated HLM streamflow during identified peak flow events at three gauges: the Cedar River at Cedar Rapids (6,492 km^2), the Iowa River at Lone Tree (4,291 km^2), and East Nishnabotna River at Red Oak (897 km^2). Visual comparison of observed and simulated streamflow reveals that Stage IV-based estimates generally match observations leading up to and during high flow events; however, even though Stage IV precipitation data is relatively accurate, Stage IV-based streamflow peaks often lag behind observed peaks, indicating a timing error in the hydrologic model. Further, Stage IV-based streamflow does not agree with observed streamflow during the recession period (e.g. late June 2008 at the Cedar River in Figure 8). In almost all high flow events illustrated in Figure 8, IMERG-based streamflow overestimates the actual streamflow, sometimes by a considerable margin. This is consistent with findings that IMERG-Early tends to overestimate extreme precipitation in this part of the world (Li et al., 2022). The STREAM simulation mean generally matches Stage IV and observations more than IMERG, and the 90% confidence interval brackets the observed streamflow at almost all timesteps during increasing and peak flows. The STREAM 90% CI is widest during periods when IMERG uncertainty is high and IMERG-based streamflow estimates correspondingly exhibit high error (e.g., the 2010 event in Figure 8). All HLM simulations exhibit a tendency to underestimate streamflow during the recession period following peak events.

The MAE of IMERG- and Stage IV-based streamflow and MCRPS of STREAM-based streamflow ensembles when estimating the peak flow, event volume, and event duration calculated for the largest ten events during the study period at each gauge are shown in Figure 9a-c. As with the MAE and MCRPS for estimating streamflow, the MAE and MCRPS for estimated peak flow and event volume increase with upstream basin area. The MCRPS of STREAM-based peak flow

estimates is substantially reduced relative to the MAE of IMERG-based peak flow at gauge locations with upstream areas greater than 1,000 km^2. The accuracy of event volume predictions does not improve as much as the accuracy of peak flow when predicted using the STREAM ensemble instead of IMERG. The absolute error in estimated peak flow, event volume, and event duration for identified events at 97 gauge stations in June 2008 and 101 gauge stations in June

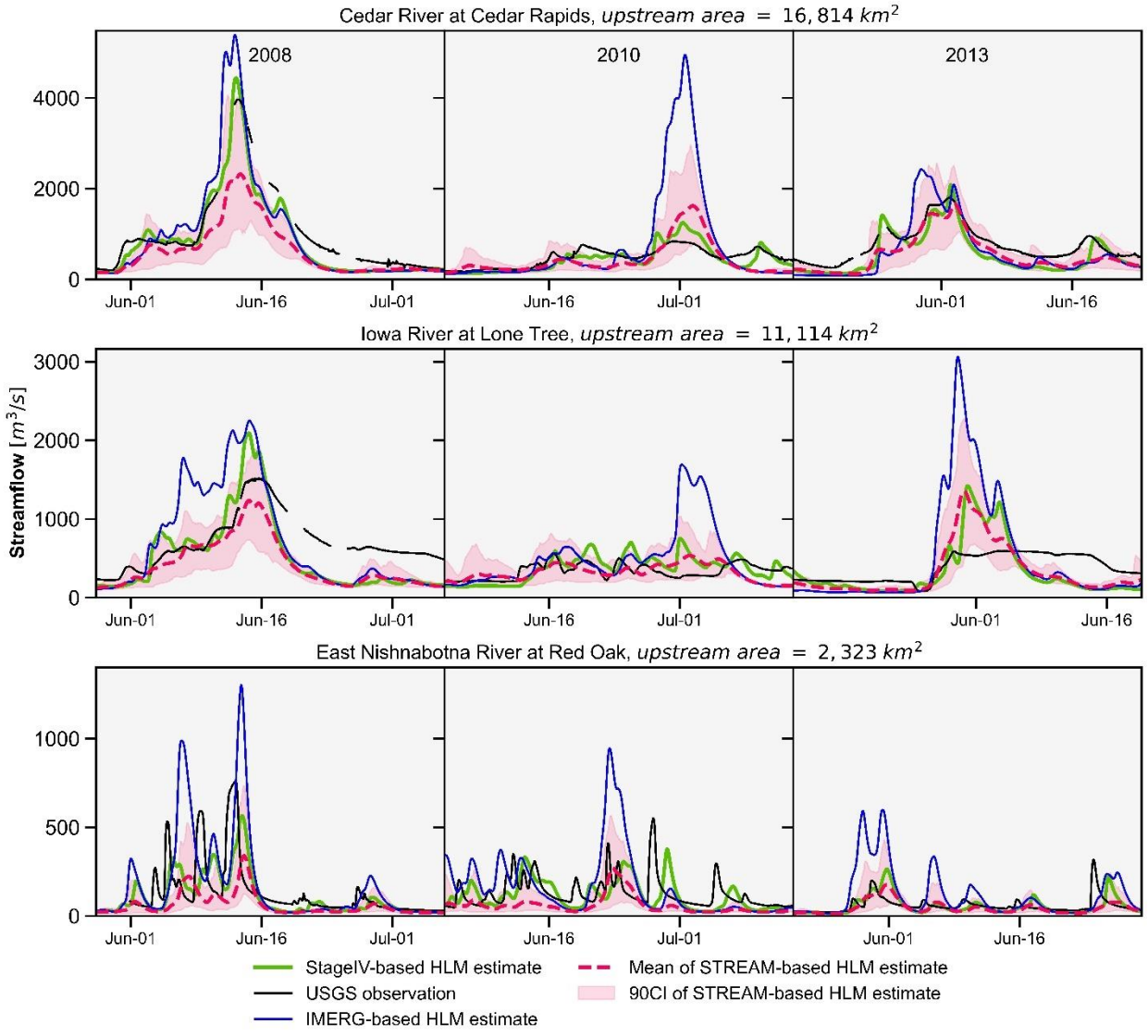


Figure 8. Example of identified peak streamflow events and performance of HLM streamflow estimates at three USGS gauge sites. Observed streamflow from USGS gauges (black) and predicted streamflow is shown for HLM results based on Stage IV (green), IMERG (blue), and the STREAM ensemble (red).

2013 are shown in Figure 9d-f and Figure 9g-i, respectively. As found in Quintero et al. (2020b), errors in peak flow estimation increase with upstream area (Figure 9d, 9g). The absolute error for STREAM simulations is shown as a 90% confidence interval, representing the range of STREAM-based estimates for peak flow, event volume, and event duration, which may be interpreted as the range of uncertainty surrounding these event metrics due to precipitation error. IMERG

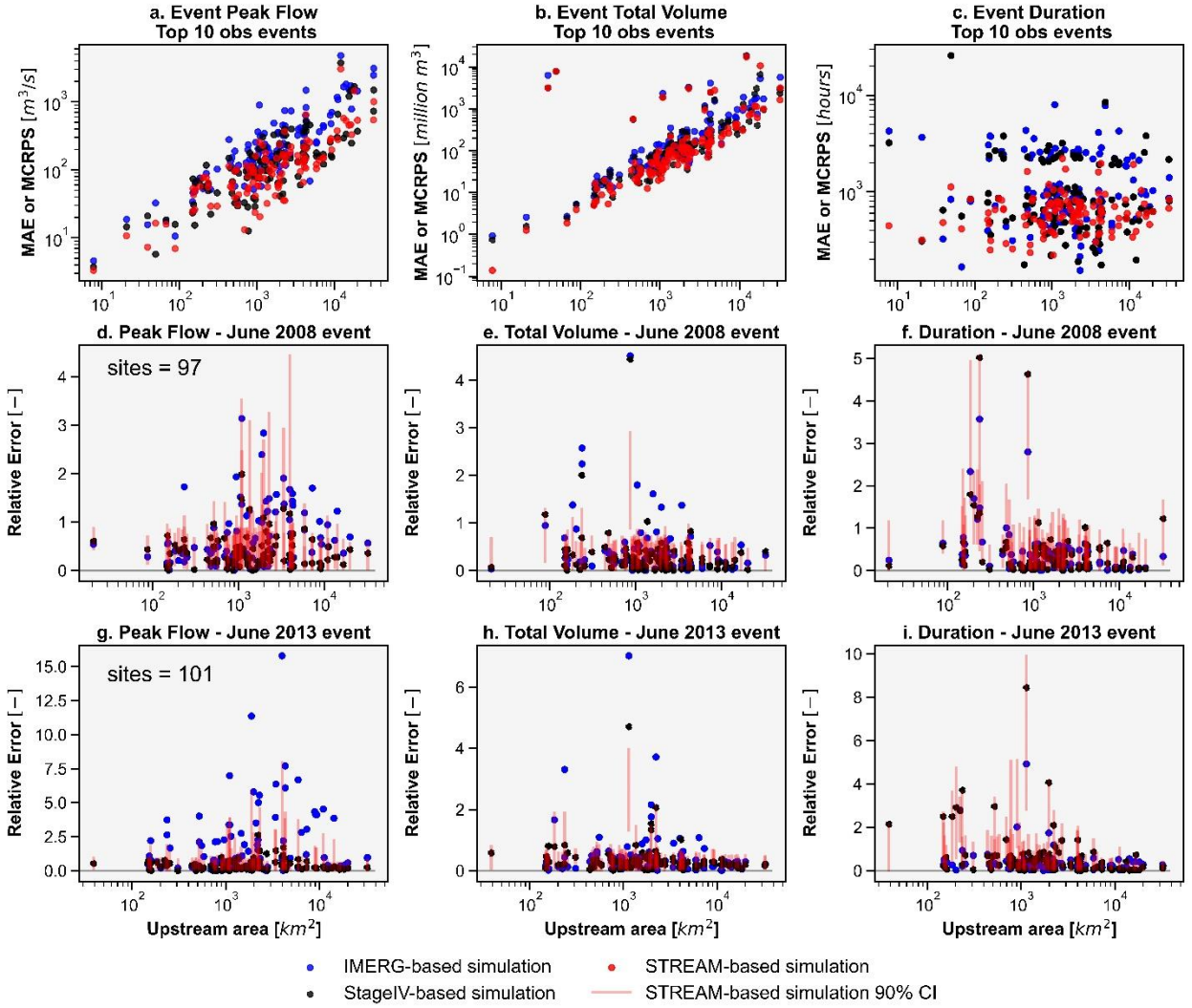


Figure 9. (Top) The MAE of IMERG-based and Stage IV-based estimates and MCRPS of STREAM-based estimates of (a) event maximum streamflow and (b) event total volume for the ten largest events at each gauge site. (Middle) The absolute error of simulated estimates of (d) peak flow, (e) event volume, and (f) event duration during the June 2008 events detected at 97 gauge sites. (Bottom) Absolute error of simulated estimates of (g) peak flow, (h) event volume, and (i) event duration during the June 2013 events detected at 101 gauge sites

consistently overestimates peak flow and volume across all sites during the 2008 event while Stage IV estimates are much closer to observed values. The 90% CI of the STREAM-based simulation consistently brackets observed peak flow and duration but fails to capture event volume in several sites with high upstream areas during the 2013 event. The performance of all simulations and range of uncertainty surrounding event duration in 2008 and 2013 are comparable across basin sizes.

4.3 Effect of precipitation uncertainty, precipitation resolution, and bias-correction

Figure 10 presents the proportion of streamflow error (here in terms of MAE) that can be attributed to different precipitation sources, sorted by upstream area. Figure 10a shows the percent reduction in MAE from IMERG-based streamflow to streamflow forced by more accurate Stage IV data (at 0.1° resolution and at native resolution). The percentage of streamflow MAE attributable to IMERG uncertainty increases with upstream area and ranges from 20 to 40% of MAE on average across basin sizes. In some basins, this percentage is as high as 60. Figure 10b

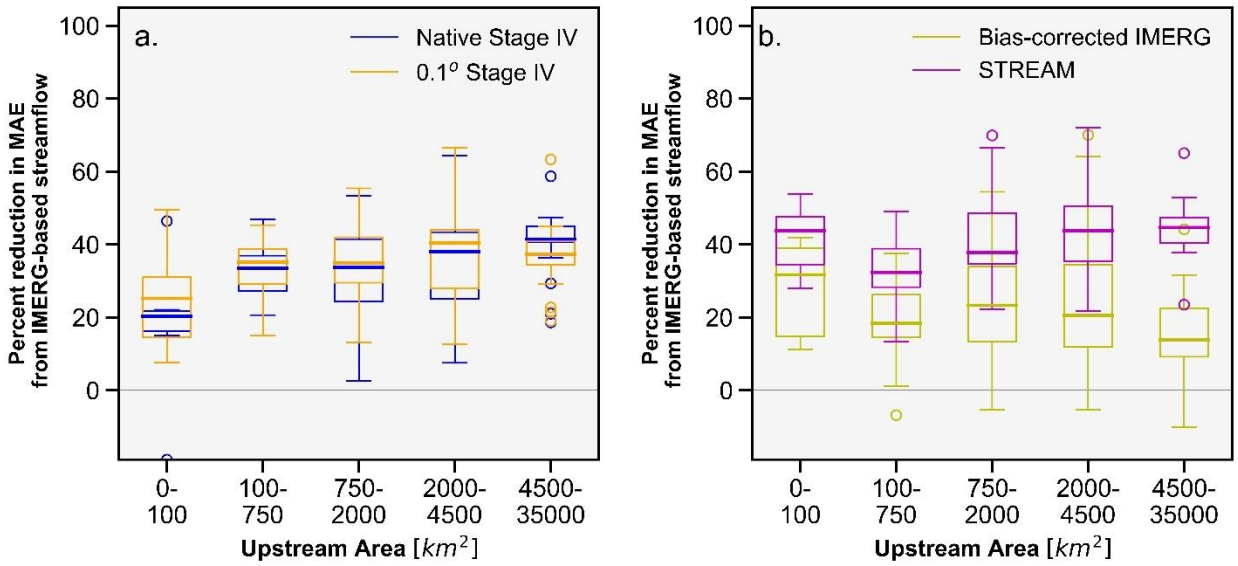


Figure 10. (a) Percent reduction in streamflow MAE from IMERG-based streamflow to native resolution and 0.1° Stage IV-based streamflow. (b) The reduction in streamflow MAE when using bias-corrected IMERG and the STREAM ensemble instead of IMERG as precipitation input to the Hillslope Link Model. The MCRPS is used to calculate the reduction in streamflow MAE for the STREAM ensemble.

compares the percent reduction of IMERG-based streamflow MAE when compared to bias-corrected IMERG (yellow; described in Section 3.3) and the STREAM ensemble simulations (magenta). The reduction in streamflow MAE is consistently higher across basin sizes when using the STREAM ensemble instead of bias-corrected IMERG to force the HLM. In some basins, using bias-corrected IMERG to predict streamflow actually increases the MAE

of predicted streamflow relative to IMERG. Unlike the STREAM-based streamflow estimates, streamflow based on bias-corrected IMERG appears to perform worse relative to IMERG as the upstream area increases at gauge sites. Although Figure 10b is calculated using all streamflow estimates, Supplemental Figure 2 shows the same analysis for the highest 10% of streamflow observations at each gauge site.

5. Discussion

5.1 Performance of HLM streamflow estimates overall and during peak flow events

Across study results, use of the STREAM ensemble to force the HLM hydrologic model improves streamflow performance to nearly the same degree as use of ground reference data, relative to the satellite multisensor product IMERG. Results that demonstrate greater relative improvement in MAE in large basins when using STREAM ensemble fields to force the HLM (Figures 6,10) correspond to results in Falck et al. (2018), which found greater reduction in random error at large basin scales in Brazil. The notably greater reduction in MCRPS relative to MAE in large basins demonstrates the importance of accounting for precipitation uncertainty in hydrologic models for basins greater than 10,000 km^2 in size (Figures 5 and 6). While the error in streamflow estimates in small basins may not greatly improve when precipitation uncertainty is accounted for, accounting for precipitation uncertainty in large basins noticeably reduces error in streamflow

estimates. That STREAM-based streamflow ensembles exhibit NRMSE and MCRPS nearly as low as that of Stage IV-based streamflow ensembles across all gauge sites overall and during peak events (Figures 5 and 9) demonstrates that STREAM effectively captures the range of true precipitation and is effective at substantially reducing the effects of precipitation uncertainty on HLM output. Notably, this is contingent on both the ability of STREAM to simulate the space-time autocorrelation structure of errors as well as the ability of the CSGD error model to provide apt uncertainty characterizations for every IMERG estimate. Improved streamflow prediction using the STREAM ensemble supports previous work demonstrating the advantages of ensemble-based, probabilistic flood forecasting (e.g., Cloke & Pappenberger, 2009; Verbunt et al., 2007).

The STREAM ensemble outperforms bias-corrected IMERG because firstly, the bias-corrected product cannot simulate extreme rain rates while it is correcting for the systematic tendency of IMERG to overestimate moderate to high precipitation events and, furthermore, it cannot simulate high rain rates over multiple pixels (e.g., autocorrelated errors) in a watershed, which leads to heavy underestimation of streamflow in instances when high rain rates are, in fact, occurring across a watershed.

It is also interesting that IMERG-based streamflow does not always fall within the STREAM ensemble range, especially during peak flows (see Figure 8, Supplemental Figure 1). This occurs when IMERG overestimates precipitation over multiple pixels and the STREAM methodology generates precipitation fields that may contain some high rain rates, or a similar area of precipitation, but that overall reduce the amount of IMERG-estimated precipitation across a region. Essentially, during process of correcting for systematic bias, simulating the range of random error, and simulating error autocorrelation structures, STREAM recognizes that such a large area of extreme rainfall observed by IMERG is statistically very unlikely to correspond to

489 such extreme ‘true’ rainfall; thus, no ensemble members are generated that predict this high
490 regional precipitation as the possible true rainfall. The resulting streamflow forced by the
491 STREAM ensemble does not bracket the IMERG-based simulation, but it is clear that STREAM
492 was correct in generally reducing precipitation rates and the STREAM ensemble replicates the
493 Stage IV simulation and observations more than IMERG (Figure 8, Supplemental Figure 1).

494 The decreasing UR and increasing CR of the STREAM-based streamflow ensemble with
495 upstream area may be due to smaller relative uncertainties (UC) in larger basins (as in Falck et al.,
496 2015; Maggioni et al., 2013). These UR and CR trends also demonstrate that the higher CR in
497 larger basins is not due to a significantly larger ensemble spread relative to the average streamflow.
498 These results are consistent with previous findings by Falck et al. (2015), which also found UR to
499 decrease with increasing upstream area in Brazil. Additionally, it could be expected that basins
500 with upstream areas < 100 km^2 perform differently than larger basins (e.g., Figure 6c) because the
501 runoff in these basins is based on a single IMERG estimate and does not need to account for the
502 autocorrelation of errors in multiple IMERG estimates across a watershed. Decreasing
503 performance of bias-corrected IMERG predictions with upstream area (Figure 9b, Supplemental
504 Figure 2b) indicates that basins in this study area are not effectively filtering out precipitation
505 random error. Otherwise, larger basin areas should correspond to better performance of streamflow
506 predictions generated by IMERG and bias-corrected IMERG. That STREAM-based predictions
507 meanwhile improve in performance with upstream area (Figure 9b, Supplemental Figure 2b)
508 indicates that (1) autocorrelation of random errors has a substantial impact on the propagation of
509 precipitation error in the HLM and (2) STREAM effectively simulates such autocorrelation
510 structures.

The containing ratio and MCRPS show consistent improvement with increasing ensemble size until approximately an ensemble size of approximately 30 (Figure 7). In small basins, fewer ensemble members are required, which is intuitive given the limited variety of potential true precipitation fields that an ensemble generates for basins covered by only a handful of IMERG pixels. However, in large basins, where there are more possible combinations of precipitation error and error autocorrelation structures, more realizations of the possible true precipitation fields are required to capture observed streamflow.

The Hillslope Link Model was designed principally to monitor high streamflow conditions that may endanger communities and infrastructure in Iowa. It is therefore not unexpected that its streamflow estimates during low flow conditions do not consistently match USGS observations (see Figure 8; Quintero, et al., 2020). Given the inaccuracy in HLM streamflow estimates during low flow conditions or recessions after peak events even when the high-quality Stage IV product is used (e.g. Figure 8) provides reasonable grounds to conclude that poor streamflow accuracy at low flows is due to structural or parameter error within HLM rather that problems with the precipitation. Visual inspection of hydrographs also reveals timing errors in predicted peak streamflow events that can be attributed to either the HLM or insufficient temporal resolution of precipitation forcing data, but not to precipitation uncertainty since even the high accuracy ground reference precipitation dataset fails to generate correct timing of peak streamflow in some instances (Figure 8, bottom left panel).

5.2 Effect of precipitation uncertainty, precipitation resolution, and bias-correction

A sizeable percentage—between 20 and 40% on average across basin sizes—of the error in predicted streamflow using IMERG data can be attributed to errors in this precipitation data

product (Figure 10a). The minimal difference in reduction in predicted streamflow MAE when using 1/24° and 0.1° resolution Stage IV demonstrates that uncertainty in precipitation data makes a larger contribution towards the error in simulated streamflow than does the resolution of the precipitation data. Coarser resolution Stage IV improves predicted streamflow accuracy slightly more that native resolution Stage IV, except in the largest basins (Figure 10a). When only the top 10% of streamflow observations are considered, however, native resolution Stage IV outperforms the coarser resolution product in all basin size categories except those less than 100 km^2 (Supplemental Figure 2). The ability of coarser resolution ground reference data to produce streamflow estimates with a slightly lower MAE than 1/24° resolution ground reference data is most likely due to sampling error; several of the USGS gauge sites for basins in this study only have available data for two or three of the years in the six years study period.

Correcting for systematic bias does not account for random error, which is a substantial component of overall IMERG and other SMP product uncertainty (Figure 10b). During all flows and particularly during the highest subset of streamflow across gauge sites, streamflow error is consistently reduced more when the STREAM ensemble is used as input to the HLM rather than the bias-corrected IMERG dataset. This demonstrates the inadequacy of bias correction as an approach to addressing precipitation uncertainty in hydrologic models like the HLM and supports similar findings in (Habib et al., 2014) and (REFERENCE) Hartke et al (2020) regarding SMP uncertainty in runoff simulations and landslide hazard applications, respectively.

6. Summary and Conclusions

Although satellite multisensor precipitation products and numerical weather model fields have made possible near real-time hydrologic modeling on a continental to global scale in recent

555 decades, streamflow estimates are hampered by the uncertainty in precipitation data from these
556 sources. Overestimation, underestimation, and incorrect detection of precipitation all lead to less
557 accurate estimates of runoff and streamflow in hydrologic models. Although hydrologic model
558 calibration may reduce the propagation of forcing data uncertainty, it is no substitute for explicitly
559 representing the uncertainty in forcing precipitation data, which includes substantial random error
560 in addition to systematic bias. However, this effort is complicated by the need to simulate the
561 autocorrelation structure of precipitation errors in order to capture basin-scale precipitation
562 uncertainty, while minimizing reliance on ground references and representing precipitation
563 uncertainty in a way that hydrologic models or other applications with deterministic inputs can
564 readily ingest.

565 In this paper, precipitation ensembles generated by the Space-Time Rainfall Error and
566 Autocorrelation Model (STREAM) are used to force the Hillslope Link Model and simulate
567 streamflow ensembles which reflect the uncertainty in precipitation forcing from IMERG. Using
568 STREAM ensembles substantially reduces the error in streamflow estimates overall and during
569 peak flow events. Results show that a substantial portion of the error between predicted and
570 observed streamflow—as much as 60%—is due to the error in IMERG. Correcting for systematic
571 bias while neglecting the range of random error is shown to be an insufficient approach to
572 addressing precipitation uncertainty and can actually worsen streamflow predictions. Results also
573 reemphasize the results of Cunha et al. (2012)—that the spatial autocorrelation of precipitation
574 errors heavily impacts the ability of basins to filter out precipitation uncertainty. Although
575 STREAM is applied to SMP data in this work and in Hartke et al. (2022), it is also applicable to
576 numerical weather prediction (NWP) model fields.

Although the error characterization in this work utilizes a ground reference product for calibration, recent work by Li et al. (2021) demonstrates a method for characterizing satellite precipitation uncertainty using dual-precipitation radar data; this will allow STREAM to generate precipitation ensembles to represent satellite precipitation uncertainty anywhere in the world without ground reference data. Future work will use a similar methodology as demonstrated in this work to account for satellite precipitation uncertainty in hydrologic modeling on a continental to global scale without reliance on any ground-based precipitation data.

7. Acknowledgements

S. H. Hartke's contributions were funded by the Grainger Wisconsin Distinguished Graduate Fellowship program. D. B. Wright's contributions were funded by NASA Precipitation Measurement Mission Grant 80NSSC19K0951. We thank the Iowa Flood Center at the University of Iowa for the use of the Hillslope Link Model and associated computing resources. We thank Dr. Zhe Li for his matplotlib formatting template.

References

Aghakouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research Atmospheres*, *116*(2), 1–11. https://doi.org/10.1029/2010JD014741

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, *17*(3), 1161–1175. https://doi.org/10.5194/hess-17-1161-2013

Alvarez-Garreton, C., Ryu, D., Western, A. W., Crow, W. T., & Robertson, D. E. (2014). The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment. *Journal of Hydrology*, *519*(PD), 2763–2774. https://doi.org/10.1016/j.jhydrol.2014.07.041

Amorim, J. da S., Viola, M. R., Junqueira, R., de Oliveira, V. A., & de Mello, C. R. (2020). Evaluation of satellite precipitation products for hydrological modeling in the brazilian cerrado biome. *Water (Switzerland)*, *12*(9). https://doi.org/10.3390/W12092571

Anjum, M. N., Ding, Y., Shangguan, D., Ahmad, I., Wajid Ijaz, M., Farid, H. U., Yagoub, Y. E., Zaman, M., & Adnan, M. (2018). *Performance evaluation of latest integrated multi-satellite retrievals for Global Precipitation Measurement (IMERG) over the northern highlands of Pakistan*. https://doi.org/10.1016/j.atmosres.2018.02.010

Asong, Z. E., Razavi, S., Wheater, H. S., & Wong, J. S. (2017). Evaluation of integrated multisatellite retrievals for GPM (IMERG) over Southern Canada against ground precipitation observations: A preliminary assessment. *Journal of Hydrometeorology*, *18*(4), 1033–1050. https://doi.org/10.1175/JHM-D-16-0187.1

Biemans, H., Hutjes, R. W. A., Kabat, P., Strengers, B. J., Gerten, D., & Rost, S. (2009). Effects of precipitation uncertainty on discharge calculations for main river basins. *Journal of Hydrometeorology*, *10*(4), 1011–1025. https://doi.org/10.1175/2008JHM1067.1

Budikova, D., Coleman, J. S. M. M., Strope, S. A., & Austin, A. (2010). Hydroclimatology of the

2008 Midwest floods. *Water Resources Research*, *46*(12). https://doi.org/10.1029/2010WR009206

Charles, S. P., Chiew, F. H. S., Potter, N. J., Zheng, H., Fu, G., & Zhang, L. (2020). Impact of downscaled rainfall biases on projected runoff changes. *Hydrology and Earth System Sciences*, *24*(6), 2981–2997. https://doi.org/10.5194/hess-24-2981-2020

Ciach, G. J., Krajewski, W. F., & Villarini, G. (2007). Product-Error-Driven Uncertainty Model for Probabilistic Quantitative Precipitation Estimation with NEXRAD Data. *Journal of Hydrometeorology*, *8*(6), 1325–1347. https://doi.org/10.1175/2007jhm814.1

Ciupak, M., Ozga-Zielinski, B., Adamowski, J., Deo, R. C., & Kochanek, K. (2019). Correcting satellite precipitation data and assimilating satellite-derived soil moisture data to generate ensemble hydrological forecasts within the HBV rainfall-runoff model. *Water (Switzerland)*, *11*(10). https://doi.org/10.3390/w11102138

Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. In *Journal of Hydrology* (Vol. 375, Issues 3–4, pp. 613–626). https://doi.org/10.1016/j.jhydrol.2009.06.005

Cunha, L. K., Mandapaka, P. V, Krajewski, W. F., Mantilla, R., & Bradley, A. A. (2012). Impact of radar-rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model. *Water Resources Research*, *48*(10). https://doi.org/10.1029/2012WR012138

Demir, I., & Krajewski, W. F. (2013). Towards an integrated Flood Information System: Centralized data access, analysis, and visualization. *Environmental Modelling and Software*, *50*, 77–84. https://doi.org/10.1016/j.envsoft.2013.08.009

Demir, I., Yildirim, E., Sermet, Y., Muhammed, &, Sit, A., & Sit, M. A. (2018). FLOODSS: Iowa flood information system as a generalized flood cyberinfrastructure. *International Journal of River Basin Management*, *16*(3), 393–400. https://doi.org/10.1080/15715124.2017.1411927

Falck, A. S., Maggioni, V., Tomasella, J., Diniz, F. L. R., Mei, Y., Beneti, C. A., Herdies, D. L., Neundorf, R., Caram, R. O., & Rodriguez, D. A. (2018). Improving the use of ground-based

radar rainfall data for monitoring and predicting floods in the Iguaçu river basin. *Journal of Hydrology*, *567*, 626–636. https://doi.org/10.1016/j.jhydrol.2018.10.046

Falck, A. S., Maggioni, V., Tomasella, J., Vila, D. A., & Diniz, F. L. R. (2015). Propagation of satellite precipitation uncertainties through a distributed hydrologic model: A case study in the Tocantins-Araguaia basin in Brazil. *Journal of Hydrology*, *527*, 943–957. https://doi.org/10.1016/j.jhydrol.2015.05.042

Falck, A. S., Tomasella, J., Diniz, F. L. R., & Maggioni, V. (2021). Applying a precipitation error model to numerical weather predictions for probabilistic flood forecasts. *Journal of Hydrology*, *598*, 126374. https://doi.org/10.1016/j.jhydrol.2021.126374

Fekete, B. M., Vörösmarty, C. J., Roads, J. O., & Willmott, C. J. (2004). Uncertainties in precipitation and their impacts on runoff estimates. *Journal of Climate*, *17*(2), 294–304. https://doi.org/10.1175/1520-0442(2004)017<0294:UIPATI>2.0.CO;2

Franz, K. J., & Hogue, T. S. (2011). Evaluating uncertainty estimates in hydrologic models: Borrowing measures from the forecast verification community. *Hydrology and Earth System Sciences*, *15*(11), 3367–3382. https://doi.org/10.5194/hess-15-3367-2011

Germann, U., & Zawadzki, I. (2002). Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology. *Monthly Weather Review*, *130*(12), 2859–2873. https://doi.org/10.1175/1520-0493(2002)130<2859:SDOTPO>2.0.CO;2

Gilewski, P. P., & Nawalany, M. (2018). Inter-comparison of Rain-Gauge, Radar, and Satellite (IMERG GPM) precipitation estimates performance for rainfall-runoff modeling in a mountainous catchment in Poland. *Water (Switzerland)*, *10*(11). https://doi.org/10.3390/w10111665

Habib, E., Haile, A. T., Sazib, N., Zhang, Y., & Rientjes, T. (2014). Effect of bias correction of satellite-rainfall estimates on runoff simulations at the source of the Upper Blue Nile. *Remote Sensing*, *6*(7), 6688–6708. https://doi.org/10.3390/rs6076688

Hartke, S. H., Wright, D. B., Kirschbaum, D. B., Stanley, T. A., & Li, Z. (2020). Incorporation of

satellite precipitation uncertainty in a landslide hazard nowcasting system. *Journal of Hydrometeorology*, *21*(8), 1741–1759. https://doi.org/10.1175/JHM-D-19-0295.1

Hartke, S. H., Wright, D. B., Li, Z., Maggioni, V., & Dalia, B. (2021). Ensemble Representation of Satellite Precipitation Uncertainty using an Uncalibrated , Nonstationary , Anisotropic Autocorrelation Model. *Water Resources Research*. https://doi.org/10.1029/2021WR031650

Hartke, S. H., Wright, D. B., Li, Z., Maggioni, V., & Dalia, B. (2022). Ensemble Representation of Satellite Precipitation Uncertainty using a Nonstationary, Anisotropic Autocorrelation Model. *Water Resources Research*, *58*(8). https://doi.org/10.1029/2021WR031650

Hartke, S. H., Wright, D. B., Li, Z., Maggioni, V., Kirschbaum, D. B., & Khan, S. (2021). Ensemble Representation of Satellite Precipitation Uncertainty using an Uncalibrated, Nonstationary, Anisotropic Autocorrelation Model. *Earth and Space Science Open Archive*. https://doi.org/10.1002/essoar.10508893.1

Hong, Y., Hsu, K. L., Moradkhani, H., & Sorooshian, S. (2006). Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. *Water Resources Research*, *42*(7), 1–15. https://doi.org/10.1029/2005WR004398

Hossain, F., & Anagnostou, E. N. (2006). A two-dimensional satellite rainfall error model. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(6), 1511–1522. https://doi.org/10.1109/TGRS.2005.863866

Huffman, G., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Tan, J., & Xie, P. (2019). NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Prepared for: Global Precipitation Measurement (GPM) National Aeronautics and Space Administration (NASA). In *Algorithm Theoretical Basis Document (ATBD) Version 06* (Issue March). https://pmm.nasa.gov/sites/default/files/imce/times_allsat.jpg

Ji, X., Li, Y., Luo, X., He, D., Guo, R., Wang, J., Bai, Y., Yue, C., & Liu, C. (2020). Evaluation of bias correction methods for APHRODITE data to improve hydrologic simulation in a large

Himalayan basin. *Atmospheric Research*, *242*(February), 104964. https://doi.org/10.1016/j.atmosres.2020.104964

Jiang, L., & Bauer-Gottwein, P. (2019). How do GPM IMERG precipitation estimates perform as hydrological model forcing? Evaluation for 300 catchments across Mainland China. *Journal of Hydrology*, *572*(March), 486–500. https://doi.org/10.1016/j.jhydrol.2019.03.042

Kirstetter, P. E., Karbalaee, N., Hsu, K., & Hong, Y. (2018). Probabilistic precipitation rate estimates with space-based infrared sensors. *Quarterly Journal of the Royal Meteorological Society*, *144*(December 2017), 191–205. https://doi.org/10.1002/qj.3243

Krajewski, W. F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., Mantilllla, R., Niemeier, J., Quintero, F., Seo, B. C., Smallll, S. J., Weber, L. J., & Young, N. C. (2017). Real-time flood forecasting and information system for the state of Iowa. *Bulletin of the American Meteorological Society*, *98*(3), 539–554. https://doi.org/10.1175/BAMS-D-15-00243.1

Li, N., Tang, G., Zhao, P., Hong, Y., Gou, Y., & Yang, K. (2016). *Statistical assessment and hydrological utility of the latest multi-satellite precipitation analysis IMERG in Ganjiang River basin*. https://doi.org/10.1016/j.atmosres.2016.07.020

Li, Z., Tang, G., Kirstetter, P., Gao, S., Li, J.-L. L. F., Wen, Y., & Hong, Y. (2022). Evaluation of GPM IMERG and its constellations in extreme events over the conterminous united states. *Journal of Hydrology*, *606*, 127357. https://doi.org/10.1016/j.jhydrol.2021.127357

Li, Z., Wright, D., Hartke, S., Kirschbaum, D., Khan, S., Maggioni, V., & Kirstetter, P.-E. (2021). Toward A Globally-Applicable Uncertainty Quantification Framework for Satellite Multisensor Precipitation Products based on GPM DPR. *Earth and Space Science Open Archive*. https://doi.org/10.1002/ESSOAR.10507263.1

Lin, Y. (2011). *GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Version 1.0*. https://data.eol.ucar.edu/dataset/21.093

Liu, Y., Brown, J., Demargne, J., & Seo, D. J. (2011). A wavelet-based approach to assessing timing errors in hydrologic predictions. *Journal of Hydrology*, *397*(3–4), 210–224.

https://doi.org/10.1016/j.jhydrol.2010.11.040

Lott, N. (1993). *Research Customer Service Group Technical Report 93-04 the Summer of 1993 : Flooding in the Midwest and Drought in the Southeast*. *93*(4).

Lowrey, M. R. K., & Yang, Z. L. (2008). Assessing the capability of a regional-scale weather model to simulate extreme precipitation patterns and flooding in central Texas. *Weather and Forecasting*, *23*(6), 1102–1126. https://doi.org/10.1175/2008WAF2006082.1

Luitel, B., Villarini, G., & Vecchi, G. A. (2018). Verification of the skill of numerical weather prediction models in forecasting rainfall from U.S. landfalling tropical cyclones. *Journal of Hydrology*, *556*, 1026–1037. https://doi.org/10.1016/j.jhydrol.2016.09.019

Maggioni, V., Reichle, R. H., & Anagnostou, E. N. (2011). The effect of satellite rainfall error modeling on soil moisture prediction uncertainty. *Journal of Hydrometeorology*, *12*(3), 413–428. https://doi.org/10.1175/2011JHM1355.1

Maggioni, V., Sapiano, M. R. P., Adler, R. F., Tian, Y., & Huffman, G. J. (2014). An Error Model for Uncertainty Quantification in High-Time-Resolution Precipitation Products. *Journal of Hydrometeorology*, *15*(3), 1274–1292. https://doi.org/10.1175/jhm-d-13-0112.1

Maggioni, V., Vergara, H. J., Anagnostou, E. N., Gourley, J. J., Hong, Y., & Stampoulis, D. (2013). Investigating the Applicability of Error Correction Ensembles of Satellite Rainfall Products in River Flow Simulations. *Journal of Hydrometeorology*, *14*(4), 1194–1211. https://doi.org/10.1175/jhm-d-12-074.1

Marc, O., Jucá Oliveira, R. A., Gosset, M., Emberson, R., & Malet, J.-P. (2022). Global assessment of the capability of satellite precipitation products to retrieve landslide-triggering extreme rainfall events. *Earth Interactions*, 1–42. https://doi.org/10.1175/ei-d-21-0022.1

Moosavi, A., Rao, V., & Sandu, A. (2021). Machine learning based algorithms for uncertainty quantification in numerical weather prediction models. *Journal of Computational Science*, *50*, 101295. https://doi.org/10.1016/j.jocs.2020.101295

Mutel, C. F. (2010). *A watershed year: anatomy of the Iowa floods of 2008*. University of Iowa

Press.

Nasrollahi, N., Aghakouchak, A., Li, J., Gao, X., Hsu, K., & Sorooshian, S. (2012). Assessing the impacts of different WRF precipitation physics in hurricane simulations. *Weather and Forecasting*, *27*(4), 1003–1016. https://doi.org/10.1175/WAF-D-10-05000.1

Nijssen, B., & Lettenmaier, D. P. (2004). Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. *Journal of Geophysical Research D: Atmospheres*, *109*(2), 1–15. https://doi.org/10.1029/2003jd003497

Nikolopoulos, E. I., Anagnostou, E. N., Hossain, F., Gebremichael, M., & Borga, M. (2010). Understanding the scale relationships of uncertainty propagation of satellite rainfall through a distributed hydrologic model. *Journal of Hydrometeorology*, *11*(2), 520–532. https://doi.org/10.1175/2009JHM1169.1

Nimmo, J. R., Perkins, K. S., Plampin, M. R., Walvoord, M. A., Ebel, B. A., & Mirus, B. B. (2021). Rapid-Response Unsaturated Zone Hydrology: Small-Scale Data, Small-Scale Theory, Big Problems. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.613564

Nogueira, M. (2020). Inter-comparison of ERA-5, ERA-interim and GPCP rainfall over the last 40 years: Process-based analysis of systematic and random differences. *Journal of Hydrology*, *583*. https://doi.org/10.1016/j.jhydrol.2020.124632

Omranian, E., Sharif, H. O., & Tavakoly, A. A. (2018). How well can Global Precipitation Measurement (GPM) capture hurricanes? Case study: Hurricane harvey. *Remote Sensing*, *10*(7). https://doi.org/10.3390/rs10071150

Pradhan, A., & Indu, J. (2021). Assessment of SM2RAIN derived and IMERG based precipitation products for hydrological simulation. *Journal of Hydrology*, *603*(PD), 127191. https://doi.org/10.1016/j.jhydrol.2021.127191

Pulkkinen, S., Nerini, D., Pérez Hortal, A. A., Velasco-Forero, C., Seed, A., Germann, U., & Foresti, L. (2019). Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geoscientific Model Development*, *12*(10), 4185–4219.

https://doi.org/10.5194/gmd-12-4185-2019

Quintero, F., Krajewski, W. F., Mantilla, R., Small, S., & Seo, B. C. (2016). A spatial-dynamical framework for evaluation of satellite rainfall products for flood prediction. *Journal of Hydrometeorology*, *17*(8), 2137–2154. https://doi.org/10.1175/JHM-D-15-0195.1

Quintero, F., Krajewski, W. F., & Rojas, M. (2020). A flood potential index for effective communication of streamflow forecasts at ungauged communities. *Journal of Hydrometeorology*, *21*(4), 807–814. https://doi.org/10.1175/JHM-D-19-0212.1

Quintero, F., Krajewski, W. F., Seo, B. C., & Mantilla, R. (2020). Improvement and evaluation of the Iowa Flood Center Hillslope Link Model (HLM) by calibration-free approach. *Journal of Hydrology*, *584*(November 2019), 124686. https://doi.org/10.1016/j.jhydrol.2020.124686

Raleigh, M. S., Lundquist, J. D., & Clark, M. P. (2015). Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrology and Earth System Sciences*, *19*(7), 3153–3179. https://doi.org/10.5194/hess-19-3153-2015

Scheuerer, M., & Hamill, T. M. (2015). Statistical Postprocessing of Ensemble Precipitation Forecasts by Fitting Censored, Shifted Gamma Distributions*. *Monthly Weather Review*. https://doi.org/10.1175/MWR-D-15-0061.1

Schreiner-McGraw, A. P., & Ajami, H. (2020). Impact of Uncertainty in Precipitation Forcing Data Sets on the Hydrologic Budget of an Integrated Hydrologic Model in Mountainous Terrain. *Water Resources Research*, *56*(12). https://doi.org/10.1029/2020WR027639

Serpetzoglou, E., Anagnostou, E. N., Papadopoulos, A., Nikolopoulos, E. I., & Maggioni, V. (2010). Error propagation of remote sensing rainfall estimates in soil moisture prediction from a land surface model. *Journal of Hydrometeorology*, *11*(3), 705–720. https://doi.org/10.1175/2009JHM1166.1

Shrestha, A., Nair, A. S., & Indu, J. (2020). Role of precipitation forcing on the uncertainty of land surface model simulated soil moisture estimates. *Journal of Hydrology*, *580*(February 2019), 124264. https://doi.org/10.1016/j.jhydrol.2019.124264

Smith, J. A., Baeck, M. L., Villarini, G., Wright, D. B., & Krajewski, W. (2013). Extreme flood response: The june 2008 flooding in Iowa. *Journal of Hydrometeorology*, *14*(6), 1810–1825. https://doi.org/10.1175/JHM-D-12-0191.1

Sperna Weiland, F. C., Vrugt, J. A., van Beek, R. L. P. H., Weerts, A. H., & Bierkens, M. F. (2015). Significant uncertainty in global scale hydrological modeling from precipitation data errors. *Journal of Hydrology*, *529*, 1095–1115. http://dx.doi.org/10.1016/j.jhydrol.2015.08.061

Széles, B., Parajka, J., Hogan, P., Silasari, R., Pavlin, L., Strauss, P., & Blöschl, G. (2020). The Added Value of Different Data Types for Calibrating and Testing a Hydrologic Model in a Small Catchment. *Water Resources Research*, *56*(10). https://doi.org/10.1029/2019WR026153

Tan, J., Huffman, G. J., Bolvin, D. T., & Nelkin, E. J. (2019). IMERG V06: Changes to the morphing algorithm. *Journal of Atmospheric and Oceanic Technology*, *36*(12), 2471–2482. https://doi.org/10.1175/JTECH-D-19-0114.1

Towler, E., & McCreight, J. L. (2021). A wavelet-based approach to streamflow event identification and modeled timing error evaluation. *Hydrology and Earth System Sciences*, *25*(5), 2599–2615. https://doi.org/10.5194/hess-25-2599-2021

Tramblay, Y., Bouvier, C., Ayral, P. A., & Marchandise, A. (2011). Impact of rainfall spatial distribution on rainfall-runoff modelling efficiency and initial soil moisture conditions estimation. *Natural Hazards and Earth System Science*, *11*(1), 157–170. https://doi.org/10.5194/nhess-11-157-2011

Tramblay, Y., Bouvier, C., Martin, C., Didon-Lescot, J. F., Todorovik, D., & Domergue, J. M. (2010). Assessment of initial soil moisture conditions for event-based rainfall-runoff modelling. *Journal of Hydrology*, *387*(3–4), 176–187. https://doi.org/10.1016/j.jhydrol.2010.04.006

Vennapusa, P. K. R., & White, D. J. (2015). Performance Assessment of Secondary-Roadway Infrastructure in Iowa after 2011 Missouri River Flooding. *Journal of Infrastructure Systems*,

21(4), 1–11. https://doi.org/10.1061/(asce)is.1943-555x.0000255

Verbunt, M., Walser, A., Gurtz, J., Montani, A., & Schär, C. (2007). Probabilistic flood forecasting with a limited-area ensemble prediction system: Selected case studies. *Journal of Hydrometeorology*, *8*(4), 897–909. https://doi.org/10.1175/JHM594.1

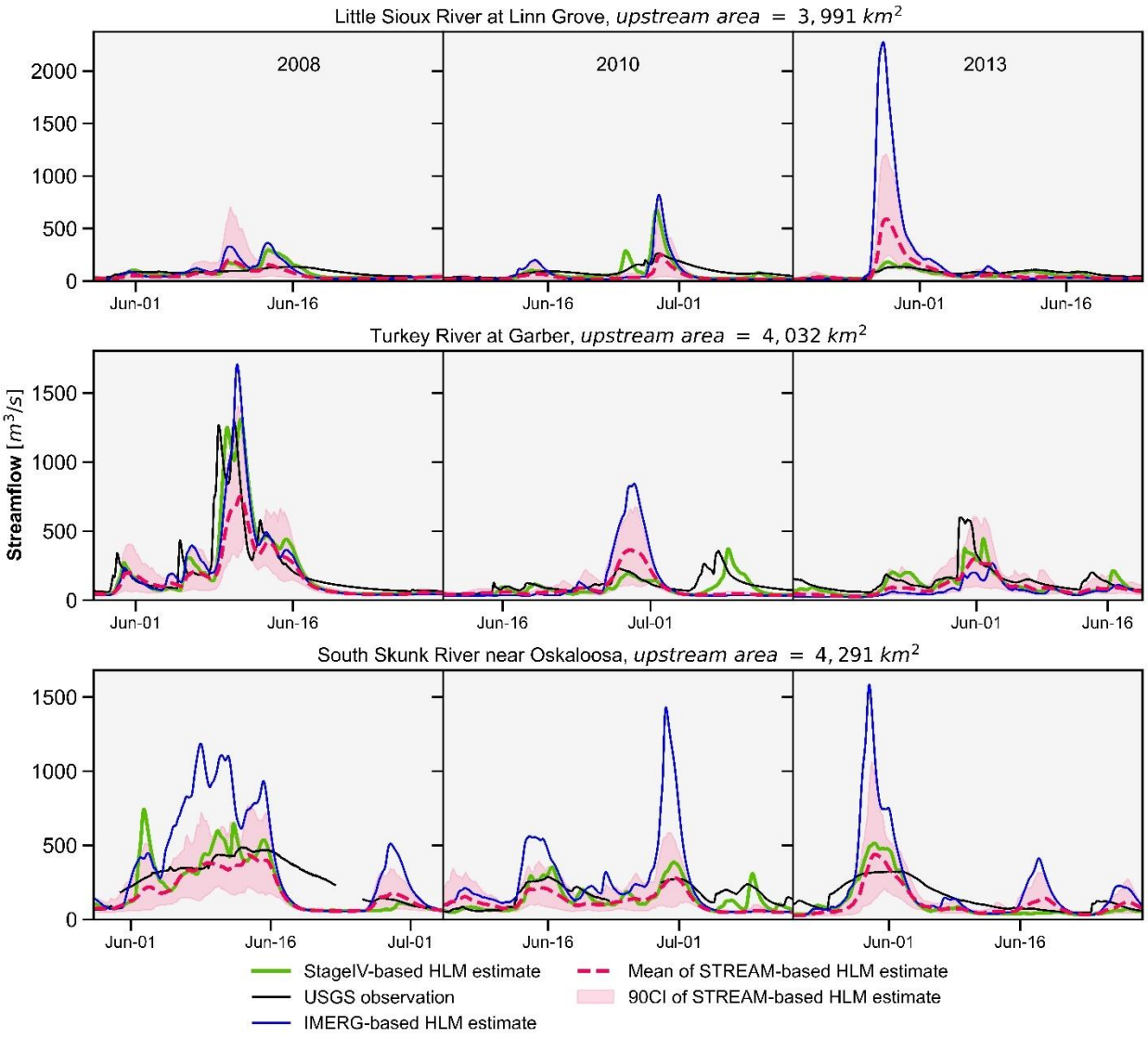
Wang, Z., Zhong, R., Lai, C., & Chen, J. (2017). Evaluation of the GPM IMERG satellite-based precipitation products and the hydrological utility. *Atmospheric Research*, *196*, 151–163. https://doi.org/10.1016/j.atmosres.2017.06.020

Wright, D. B., Kirschbaum, D. B., & Yatheendradas, S. (2017). Satellite Precipitation Characterization, Error Modeling, and Error Correction Using Censored Shifted Gamma Distributions. *Journal of Hydrometeorology*, *18*(10), 2801–2815. https://doi.org/10.1175/JHM-D-17-0060.1

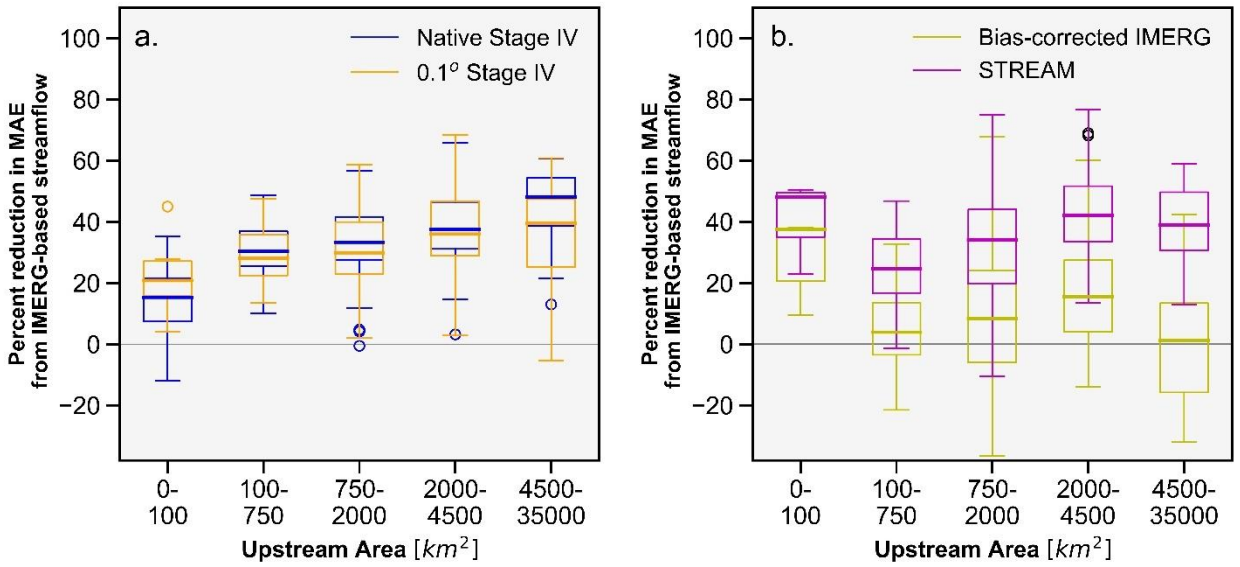
Wu, H., Adler, R. F., Tian, Y., Huffman, G. J., Li, H., & Wang, J. (2014). Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resources Research*, *50*(3), 2693–2717. https://doi.org/10.1002/2013WR014710

Xiong, L., & O'Connor, K. M. (2008). An empirical method to improve the prediction limits of the GLUE methodology in rainfall-runoff modeling. *Journal of Hydrology*, *349*(1–2), 115–124. https://doi.org/10.1016/j.jhydrol.2007.10.029

Supplemental Material



Supplemental Figure 1. As in Figure 8, example of identified peak streamflow events and performance of HLM streamflow estimates at three additional USGS gauge sites. Observed streamflow from USGS gauges (black) and predicted streamflow is shown for HLM results based on Stage IV (green), IMERG (blue), and the STREAM ensemble (red).



Supplemental Figure 2. As in Figure 10, but for top 10% of USGS streamflow observations. (a) Percent reduction in streamflow MAE from IMERG-based streamflow to native resolution and 0.1° Stage IV-based streamflow. This represents the percentage of streamflow MAE that is attributable to IMERG uncertainty and demonstrates the relatively small portion of streamflow MAE that is attributable to the resolution difference in precipitation data. (b) The reduction in streamflow MAE when using bias-corrected IMERG and the STREAM ensemble instead of IMERG as precipitation input to the Hillslope Link Model. The MCRPS is used to calculate the reduction in streamflow MAE for the STREAM ensemble.

852